

Review

Novel and emerging mutations of SARS-CoV-2: Biomedical implications



Elmira Mohammadi^{a,b}, Fatemeh Shafiee^c, Kiana Shahzamani^d, Mohammad Mehdi Ranjbar^e, Abbas Alibakhshi^f, Shahrzad Ahangarzadeh^{g,*}, Leila Beikmohammadi^{h,i}, Laleh Shariati^{h,i}, Soodeh Hooshmandi^k, Behrooz Ataei^g, Shaghayegh Haghjooy Javanmard^j

^a Applied Physiology Research Center, Cardiovascular Research Institute, Department of Physiology, Isfahan University of Medical Sciences, Isfahan, Iran

^b Core Research Facilities, Isfahan University of Medical Sciences, Isfahan, Iran

^c Department of Pharmaceutical Biotechnology, School of Pharmacy and Pharmaceutical Sciences, Isfahan University of Medical Sciences, Isfahan, Iran

^d Isfahan Gastroenterology and Hepatology Research Center (IGHRC), Isfahan University of medical sciences, Isfahan, Iran

^e Razi Vaccine and Serum Research Institute, Agricultural Research, Education, and Extension Organization (AREEO), Karaj, Iran

^f Molecular Medicine Research Center, Hamadan University of Medical Sciences, Hamadan, Iran

^g Infectious Diseases and Tropical Medicine Research Center, Isfahan University of Medical Sciences, Isfahan, Iran

^h Department of Biochemistry, Erasmus University Medical Center, Rotterdam, The Netherlands

ⁱ Stem Cell and Regenerative Medicine Center of Excellence, Tehran University of Medical Sciences, 14155-6559 Tehran, Iran

^j Nosocomial Infection Research Center, Isfahan University of Medical Sciences, Isfahan, Iran

^k Rajaie Cardiovascular Medical and Research Center, Iran University of Medical Sciences, Tehran, Iran

ARTICLE INFO

Keywords:

SARS-CoV-2

COVID-19

Mutation

Genetic variation

Spike protein

ABSTRACT

Coronavirus disease-19 (COVID-19) is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The SARS-CoV-2 virus strains has geographical diversity associated with diverse severity, mortality rate, and response to treatment that were characterized using phylogenetic network analysis of SARS-CoV-2 genomes. Although, there is no explicit and integrative explanation for these variations, the genetic arrangement, and stability of SARS-CoV-2 are basic contributing factors to its virulence and pathogenesis. Hence, understanding these features can be used to predict the future transmission dynamics of SARS-CoV-2 infection, drug development, and vaccine. In this review, we discuss the most recent findings on the mutations in the SARS-CoV-2, which provide valuable information on the genetic diversity of SARS-CoV-2, especially for DNA-based diagnosis, antivirals, and vaccine development for COVID-19.

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is an enveloped RNA virus, belongs to the *Betacoronavirus* genus. This genus includes zoonotic RNA viruses that have led to recent important epidemics: SARS (Severe Acute Respiratory Syndrome) in 2002 and MERS (Middle East respiratory syndrome) in 2012 [1]. Worst of all, in 2019, a pandemic by SARS-CoV-2 emerged in Wuhan city with acute pneumonia symptoms, which spread fast globally. The rapid dissemination and the worldwide outbreak of coronavirus disease 2019 (COVID-19) have caused to notify essential questions about the viral population's adaptation and evolution caused by mutations, recombination, and deletions. The viral genome mutations related to different agents and factors such as viral replication enzymes, host enzymes, recombination events, spontaneous nucleic acid damages, and particular genetic elements are

liable for generating new variants [2].

Remarkably, genetic differences in RNA viruses' genome occurred while being mainly distributed by an individual outbreak as a natural by-product of viral replication [3]. However, in the case of SARS-CoV-2, fewer mutations have been reported than most RNA viruses because they possess an enzyme that can correct some errors of replication. As the first line of defense against a broad range of pathogens, especially viruses, innate immune response, has raised this evolution and adaptation [4]. Viral adaptation brings a balance between genome variability and the integrity of genetic information, and the rate of mutation in RNA virus contributes to this adaptation [5,6]. Several reports have demonstrated that the deletions in the viral genome of SARS-CoV-2; often generate deletion variants of accessory and non-structural proteins, which probably directly affect viral virulence [7–9]. Upon genomic analysis, SARS-CoV-2 has various variants categorized into three types

* Corresponding author.

E-mail addresses: sh.ahangarzadeh@idrc.mui.ac.ir, shahrzadahangar@yahoo.com (S. Ahangarzadeh).

<https://doi.org/10.1016/j.bioph.2021.111599>

Received 31 January 2021; Received in revised form 18 March 2021; Accepted 27 March 2021

Available online 23 April 2021

0753-3322/© 2021 Published by Elsevier Masson SAS. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

A, B, and C. Although these SARS-CoV-2 types possess different gene mutations, these mutations' effect is debat [10]. Moreover, geographical propagations may lead to different genetic strains of SARS-CoV-2 from various locations [11].

In this review, we aim to discuss the most recent findings on the mutations in the SARS-CoV-2, which may be valuable information for the genetic diversity and genetic etiologies in the virus and, and for diagnostic and therapeutic strategies of SARS-CoV-2 especially for DNA-based diagnostic, drug design and vaccine development against COVID-19.

2. Evolution of viruses and mechanisms of viral mutation

Genetic diversity depends on various parameters. Among these, the mutation rate is of great importance because of its special place as a source of genetic variation [2,12]. In fact, mutations are the building blocks of most evolutions [12]. Mutation rates among viruses are very high. Most differences are present between DNA and RNA viruses (Table 1). RNA viruses mutate more quickly than DNA viruses. Studies showed that mutation rate ranges between 10 and 8–10–6 for DNA viruses and 10–6–10–4 for RNA viruses [13]. Eventually, for an unknown reason, single-stranded viruses tend to mutate faster than double-stranded ones; resulting in that mutation rates of single-stranded DNA viruses be incomparable with double-stranded RNA viruses [13]. It seems that there are no differences between single- or double-stranded RNA viruses. One possible explanation for this event is that single-stranded genomes are more sensitive to oxidative deamination and other kinds of chemical damage. Differences among single- or double-stranded RNA viruses may be due to their access to post-replicative repair [2].

Genome size seems to have a negative correlation with mutation rate. The high mutation rate of RNA viruses is related to the short length of the genome. Some DNA viruses with larger genome sizes code a DNA repair protein. RNA-dependent RNA polymerase (RdRp) is the other parameter for a higher mutation rate of RNA viruses. Unlike many polymerases involved in the replication of DNA viruses, RdRp does not have a proofreading activity. This enzyme is incapable of correcting

mistakes occurring during viral replication because it lacks a 3' exonuclease domain. Retroviruses reverse transcriptase (RT) enzyme has the same property that permits them to mutate and evolve very fast [2]. Unlike most RNA viruses, the order of Nidovirals including, Roniviruses, Toroviruses, and Coronaviruses, have an RdRp independent proof-reading activity. They encode RdRp containing a 3' exonuclease region and thus show lower mutation rates. This feature seems to be a critical factor in describing how these viruses possess larger genomes (26 kb) in comparison with other RNA viruses.

It should be noted that besides viral factors affecting the mutation rate of the virus, host encoded factors such as Double-strand RNA-dependent adenosine deaminases (ADARs), Apolipoprotein B mRNA editing catalytic polypeptide-like enzymes (APOBEC), and uracil DNA glycosylases (UNG) influence deeply on viral mutation rate [12].

Finally, the viral mutation rate is determined by multiple viral and host mechanisms including proofreading activity, environmental changes, genome size, polymerase fidelity, replication mechanisms, post replicative repair, sequence background, template secondary structure, spontaneous nucleic acid damage, editing by host-encoded deaminases, imbalances in nucleotide pools [2].

3. Genetic material of SARS-CoV-2

Wu et al. reported the first genomic sequence of SARS-CoV-2 from a worker of the Wuhan market on 26 December 2019 [46]. The accession number of this sequence in the National Center for Biotechnology Information (NCBI) GenBank [47] has been provided since January 2020 (NC_045512). The arrangement of this positive single-stranded RNA genome sequence is 5'-untranslated regions (UTR)- non-structural region (replicase complex[ORF1ab])-structural region (Spike-Envelope-Membrane-Nucleocapsid proteins)-3' UTR (Fig. 1). The non-structural region consists of replicase polyproteins, which are essential for replicating and transcribing the viral genome. Sixteen non-structural genes in this region nsp1–16 [48] encode proteins needed for the virus's biochemical and molecular functions in host cells. Nsp1 or host shutoff factor, which suppresses innate immune in the host, binds to the human ribosomal complexes in the 40 S subunit, consisting of the 43 S pre-initiation

Table 1
The viral mutation rate in human infecting viruses.

Class	Virus	Genome size (kb)	Average mutation rate	References
ss (+) RNA	Human Rhinovirus 14	7.13	6.9×10^{-5}	[14,15]
	Poliovirus 1	7.44	9.0×10^{-5}	[16–18]
	Coxsackie B virus	7.4	4.76×10^{-3}	[19]
	Human Norovirus G1	7.65	1.5×10^{-4}	[20]
	Hepatitis C virus	9.65	3.8×10^{-5}	[21,22]
	Zika virus	10	6.76×10^{-4}	[23]
	Dengue virus	10.7	7.17×10^{-4}	[24]
	Human coronavirus_229E	27.3	3.28×10^{-4}	[25,26]
	SARS coronavirus	29.7	2.80×10^{-3}	[27]
	Vesicular stomatitis virus	11.2	3.7×10^{-5}	[24,28]
ss (-) RNA	Rabies virus	11.9	3.97×10^{-4}	[29]
	Influenza A virus	13.6	2.5×10^{-5}	[30,31]
	Human Parapneumo virus	13.34	7.12×10^{-4}	[32]
	Mumps Virus	15.38	9.17×10^{-4}	[33]
	Measles virus	15.9	3.5×10^{-5}	[34]
	Respiratory Syncytial Virus	15.19	2.31×10^{-3}	[35]
	CCHFV	19.15	9.87×10^{-5}	[36]
	Human rotavirus A	18.56	1.76×10^{-3}	[37]
	Retro (Reverse transcribing)	3.22	3.21×10^{-4}	[38]
	Human T Cell Leukemia virus	8.50	1.6×10^{-5}	[39]
ds DNA	HIV-1	9.18	6.3×10^{-5}	[40,41]
	Rous sarcoma virus	9.40	1.4×10^{-4}	[38]
	Human Parvovirus_B19	5.59	1.55×10^{-4}	[42]
	JC Polyomavirus	5.13	1.7×10^{-5}	[43]
	Human Adenovirus 5	35.9	1.31×10^{-7}	[2]
	Herpes simplex virus type1	152	5.9×10^{-8}	[44]
	Human cytomegalovirus	235	2.0×10^{-7}	[45]

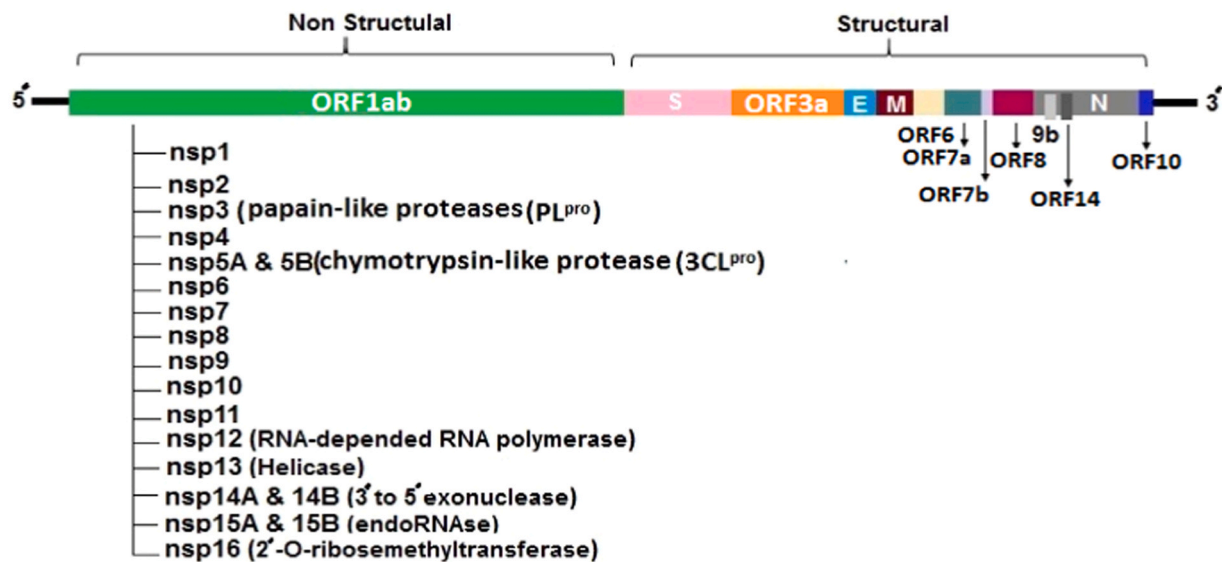


Fig. 1. Schematic structure of SARS-CoV-2 genome.

complex and the non-translating 80S ribosome [49]. Nsp2 and S protein have shown a significant stimulatory effect on the type-I interferon (IFN) induction [50]. Two proteases are encoded in the SARS-CoV-2 genome. One of them is a papain-like protease (PL^{pro}) encoded in nsp3, which cleaves nsp1, nsp2, and nsp3. The other one is 3-chymotrypsin-like “main” protease (3CL^{pro}) encoded in nsp5, which processes 13 other non-structural proteins after production [51]. It is indicated that nsp7 and nsp13 are related to T-cell immune response [52]. Nsps 12–16, encoding RNA-dependent RNA polymerase (RdRp), helicase, 3'-to-5' exonuclease; endoRNAse and 2'-O-ribose methyltransferase, respectively are linked with virus replication and transcription.

The structural region is another part of the SARS-CoV-2 genome that encodes four major structural proteins (Spike (S), Membrane (M), Envelope (E), and Nucleocapsid (N)). S protein binds to receptor (angiotensin-converting enzyme 2 (ACE2)) and fuses to the host cell membrane [53]. After S, ORF3a is located in the SARS-CoV-2 genome, which encodes an essential accessory protein. This protein can induce apoptosis, but the pro-apoptotic activity of ORF3a in SARS-CoV-2 is relatively weaker than SARS-CoV ORF3. It is probably associated with mild symptoms or asymptomatic during the early stages, which leads to spread more widely for this virus [54].

The E protein is the most mysterious structural protein among the four once, which plays a crucial role in the ER-Golgi localization and tight junction disruption [55]. M protein is the most abundant structural protein in coronaviruses that span virion membrane and along E protein have an important role in virus assembling. ORF6, ORF8, and nucleocapsid (N) proteins are other structural proteins that are inhibitors of the interferon type 1 signaling pathway. This interferon is a critical factor in the host's innate immune for the antiviral response [56,57].

4. Types of variation in SARS-CoV-2

Since the inception of the COVID-19 pandemic, exploration of SARS-CoV-2 genetic variation has been a notable subject globally, promoting the development of vaccines and diagnostics [58]. Several studies using various genomic sequences have performed the phylogenetic analysis of SARS-CoV-2 around the world [58–61]. Koyama et al. analyzed 10022 SARS-CoV-2 genomes from 68 countries. They detected 5775 distinct variants include two in-frame insertions, 11 frame-shift deletions, 36 stop-gained variants, 66 non-coding insertions, 100 in-frame deletions, 142 non-coding deletions, 484 non-coding regions mutations, 1965 synonymous mutations, and 2969 missense mutations. They finally

reported six major clades consisting of basal, D614G (the largest clade), L84S, L3606F, D448 deletion, and G392D, also 14 subclades [60].

A comprehensive study analyzed 12343 SARS-CoV-2 genome sequences isolated from around the world. Also, an investigation was performed on variants' correlation with the fatality rates in several countries. By 16 common amino acid mutations in hierarchical clustering, 28 countries were classified into three clusters [62]. Cluster 1 was China, Korea, Japan, India, and Singapore. Cluster 2 includes England, Iceland, Netherlands, Greece, Portugal, Brazil, Italy, Switzerland, Belgium, Hungary, and cluster 3 contains the USA, Taiwan, Australia, Canada, Thailand, Spain, Denmark, Congo, Germany, Sweden, Finland, France, and Luxembourg. Final analyses of variant frequency showed that the frequency of ORF1ab P4715L, S 614 G, and N 203 K/204 R variants was higher in cluster 2. The frequency of N 13 L and ORF1ab 3606 F was higher in cluster 1. According to the reported S protein 614 G variant data and ORF1ab 4715 L, as its highly linked variant, it was considerably associated with fatality rates in all 28 countries also 17 states from the United States [59].

In a recent study, Yellapu et al. analyzed 540 SARS-CoV-2 genomes from 20 different countries. They reported five distinct clusters. According to their results, 246 genomes among all 540 genomes indicated amino acid variation in one of the four structural genes. A high rate of amino acid variation was seen in spike protein. They reported that 202 genomes indicated 34 types of changes in amino acid consist of D1259H, N1178D, D1168H, A1078V, S940F, D936Y, A930V, F797C, Q675H, H655Y, D614G, A570V, A520S, H519Q, V483A, G476S, R408I, A348T, A344S, S248R, T240I, S221W, G181V, F157L, D111N, E96D, N74K, L54F, S50L, H49Y, T29I, Y28N, A27V, L5F. The D614G mutation was reported as the most frequent variation (in 160 genomes). Also, they reported 25 types of variation in N structural protein among 65 genomes include P344S, D343V, S327L, Q289H, T271I, G238C, S232T, G215S, A207G, T205I, G204R, R203K, S202N, S197L, S194L, R185C, D128Y, P46S, A35T, S23T, P14L, P13L, P6T, N4D, D3Y. In this group, R203K was reported as the most frequent (in 21 genomes). For protein M, A2S, V70I, T175M were reported in 6 genomes, and EL37R and P71L were reported just in 3 genomes [63].

According to another study, the D614G variant was reported as a high frequent mutation in different European countries, also Turkey and Iran [64–67]. Studies have shown that D614G in S protein is dominant in pandemic and also is related to more infectivity, transmission, and higher viral load [68]. Some studies exhibit that patients infected with the spike 614 G variant show higher mortality or clinical severity, but

614 G is associated with higher viral load especially in younger patients [69]. Other studies show that this mutation results in an open conformation of the spike receptor-binding domain (RBD), and increase binding affinity to ACE2 and fusion to the host cell, which in turn leads to an increase in SARS-CoV-2 transmissibility and infectivity. Moreover, lower ACE2 expression that is reported in European populations, North American, and African than that of Asian populations is associated with enhanced transmission efficiency. This indicates a positive selection for the D614G mutation [70]. However, this mutation is fortunately located outside of the RBD, so it probably does not affect vaccine design in the currently viral lineage [67]. Some mutations in the receptor binding domain such as V367F and D364Y seem that can improve the stability of the spike protein structure and lead to more efficient binding to the ACE receptor [71]. Another mutation in spike protein, V483A, to be noticeably resistant to monoclonal antibodies [63]. On 14 December 2020, a new variant of SARS-CoV-2 was reported from the United Kingdom and named SARS-CoV-2 VOC 202012/01. This variant is consisting of 23 nucleotide substitutions. There is no detected phylogenetically relation between this variant and the SARS-CoV-2 virus flowing in the U. K. [72]. This "English variant" is including multiple mutations in spike protein (deletion 69–70, deletion 144, substitutions S982A, T716I, P681H, D614G, A570D, N501Y, and also D1118H) in addition to the other genomic regions mutations [73]. Among spike mutations, N501Y is one of the six important residues responsible for the interaction of RBD with ACE2, so this is a major worry and has been related to enlarged infectivity and virulence in mouse models [74,75]. In addition, the E484K mutation in spike protein is an important mutation in the English variant and linked with the escape from the neutralizing antibodies against SARS-CoV-2 [76]. Another variant of SARS-CoV-2 named 20 H/501Y.V2 emerged in South Africa which was independent of the English variant. Three particular spike protein mutations in this variant are N501Y, E484K, and K417N. These three mutations were also detected in another new variant in Brazil (named 501Y.V3) [77]. Other spike mutations with less concern in South Africa variant are A701V, R246I, D215G, D80A, and L18F [78]. The other detected mutations in this variant are K165N, P71L, T205I, and SGF 3675–3677 deletion [79].

Besides mutation in structural proteins, the mutation in RdRp which is responsible for replication (such as P4715L) can be effective for virus proliferation [60]. On the other hand, the virus mutagenic capability is related to the fidelity of RdRp, which shows the significance of mutation in RdRp [1].

5. Variant analysis of SARS-CoV-2 genome

RNA viruses such as SARS-CoV-2 usually show a high mutation rate that may be due to the deficiency of proofreading abilities generally by an RNA-dependent RNA polymerase (RdRp) [80]. There are several studies on the variant analysis of the SARS-CoV-2 genome worldwide, within the year of the virus pandemic. For the whole genome of SARS-CoV-2 with 29,870 nucleotides and 9744 amino acid residues, large amounts of mutations were reported from December 2019 until now in 5' and 3' UTR, intergenic regions, and the coding sequences [81].

In a large analysis of variants on 48,635 SARS-CoV-2 genomes, a total of 353,341 mutation events were recorded [82]. Another fact in the mentioned project is that although 256 samples, mostly from Asia, did not have any variation in comparison to the NC_045512.2 Wuhan reference genome, 48,379 genome samples were specified with at least one mutation. The number of mutations is relatively low, 7.23 on average per sample. However, few samples harboring more than 15 mutations in the whole genome [82].

Overall, for 5' and 3' UTRs, it was reported that mutations occur in high frequency. For example, in the study of Rouchka, et al., surveyed 1043 filtered sequences, about 229 and 152 mutations were reported for 5' and 3' UTRs, respectively [83]. These statics from another study conducted by Rafiul Islam investigated a total of 2492 genome sequences and generally reported 1516 nucleotide-level variations, is 105

and 158 cases, respectively [84]. The final study, in this case, was a variant analysis of 10,022 genomes and representation of 260, and 224 variants for two parts of the genome from the total 65,776 variants [60]. However, because these parts of the virus genome are non-coding sequences, their final impact was not properly understood. But it was clear that 29742G > T, in the 3' UTR stem-loop II-like motif (s2m), is important in involving host transcriptional machinery and virus replication [84]. The mentioned mutation was detected in >1% within SARS-CoV-2 isolates. Furthermore, two deletion mutations were detected in the 3' UTR from a total of 2492 surveyed sequences [84].

Although the mutations of the SARS-CoV-2 genome are very common, there are some variants with higher frequency. For example, Wang et al. [85], identified ten high-frequency mutations within 108 genome isolates and used them to classify SARSCoV-2 into five main groups. In the other study conducted by Yang, 14 prevalent mutations with a mutation frequency of > 0.1 were identified [82] which will be discussed later.

The mutations in the coding sequences of SARS-CoV-2 were more investigated because of their effects on the structure or function of the protein and the higher length of the protein-coding sequence in the genome. Some of the variations in the studied genome are very routine, led to classify the genome sequence as clades. In one study, totally, 10,022 SARS-Cov-2 genomes were analyzed and it was shown that exactly 65,776 variants occurred in the surveyed genomes which 2969, and 1965 of them were missense and synonymous mutations, respectively [60]. Due to the longer size of the ORF1ab, 1905 missense variants (about 64%) were found in it. Again, in the ORF1ab, NSP3, was more studied for its variants and it was revealed that the two most frequent variants were the synonymous variant 3037 C > T (6334 samples), and A58T (2891 G > A, in the level genome) missense variant with 159 samples. On the other hand, in ORF1ab, the 14408 C > T (P4715L) variant of the RdRp gene in 6319 samples was the other most prevalent missense mutation. For the S protein, the most common missense mutation was 23403A > G (D614G) with 6294 total samples (68 countries) and assumed as the most common point mutation with 36,500 positive cases from a total of 48,635 samples [82].

In ORF3a, 25563 G > T (Q57H in the proteome level) with 2893 samples (from total 10,022 full-length genome), was the most prevalent missense variant. Finally, for ORF1ab, NSP2, 2442 positive samples for 1059C > T (T265I) was detected. Most samples containing 23403 A > G, also display the non-coding variant 241C > T, the synonymous variant 3037C > T and ORF1ab P4715L [60] (Table 2).

5.1. Classification of routine missense variants in clades

Based on the most occurred missense mutations, the variants were classified into some clades. Koyama, et al., provided a division of total clades with several sub-clades, for the first time. The first clade named Basal clade originated from China in Dec 2019, and the number of cases belonging to this clade is gradually decreasing. D614G (also called G clade), was the name of the second clade in which all genomes contain a missense mutation in the S protein, 23403A > G. The amino acid in this position of S protein, play an important role in the entrance of SARS-CoV-2 into the host cell via the ACE2 human receptor [60]. It was established that three other missense mutations including 14408 C > T, 241 C > T, and 3037 C > T, show a similar frequency with 23403A > G and almost always co-occurring in the same genomes and are belonging to the G clade. Within the D614G clade, D614G/Q57H/T265I subclade (also named as GH clade) forms the largest subclade with 2391 from 10,022 samples [60].

The third-largest clade was L84S clade (S clade), with a missense mutation in ORF8 (28144T > C) and with the most similarity to the basal clade. This variant originated from the travelers of Wuhan. L3606F was the other presented clade with a mutation in the NSP6 of ORF1ab in approximately 1070 samples and had the most frequency in China [60].

One of the introduced clades contains a deletion mutation in the

Table 2

An overview of the most variants of SARS-CoV-2 genomes.

Genomic change	Type of mutation	Gene/protein	Amino acid change	Final impact	Reference
241C > T	5' UTR	–	–	Important in transcription and viral packaging [17]	[1,6]
1059C > T	Missense	NSP2	T265I	Increasing the protein stability [18]	[6,10,19]
2891G > A	Missense	ORF1ab/NSP3	A58T	important role in viral replication [20]	[6,21]
3037C > T	Synonymous	ORF1ab/NSP3	F106F/F924F	no obvious effects on critical interaction to Nucleocapsid protein [22]	[1,3,6]
8782C > T	Synonymous	NSP4	S2839S	No obvious effects in infection ability and transmissibility	[6,3,10,23]
14,408C > T	Missense	ORF1ab/RdRp	P4715L/P323L	altered interaction with other components of the replication/transcription machinery or with the RNA template [24]	[1,3,6,10]
23403A > G	Missense	S	D614G	increased transmissibility for viruses [25]	[1,2,3,6]
25563G > T	Missense	ORF3a	Q57H	the binding affinity of Q57H Orf3a-S complex showed the greatest increase with S protein [26]	[6,10,15]
29742G > T	3' UTR	–	–	viral replication and recruitment of host transcriptional machinery [5]	[5]
28813G > T	Missense	S	K417N	enhancement of the binding of RBD to ACE2 (L)	[94,95]
23012G > A	Missense	S	E484K	improve escape from immune system antibodies (I)	[76,96]
23063A > T	Missense	S	N501Y	Viral infectivity and virulence (C,D)	[74,76]

NSP2 gene as an in-frame deletion led to exclude an Aspartic acid in the NSP2 protein and named as D448del. The prevalence of this clade worldwide was about 2.5%, worldwide [60].

G392D was the last clade with the least number of samples contains a mutation in the position of 1440 G > A of the NSP2 gene which has shown more abundance in Germany [60].

5.2. Transition versus transversion mutations

Generally, the most common nucleotide change in the SARS-CoV-2 genome, is C > T (as a transition mutation), accounting for 55.1% of all observed viral mutations, worldwide [82]. Surprisingly, for mutations that occurred in the SARS-CoV-2 genome, the transition versus transversion ratio was calculated as about 7:3. C > T transitions might be mediated by cytosine deaminases [60]. After that, the A > G especially in Africa, Europe, and the Americans has shown a rate of 14.8%. The third most common event worldwide, G > T, which is the most common transversion showed a rate of 12.0% (42,408 occurrences) may occur as a result of an oxo-guanine lesion from reactive oxygen species [82].

5.3. Very common mutations in the SARS-CoV-2 genome

As a summary for the most prevalent variants of the SARS-CoV-2 genome, it can be said that 28144T > C, in the ORF8, lead to a missense mutation as L84S, was recognized as a common mutation in the virus genome [61]. This mutation was reported in another study with frequency as 1669 samples from a total of 10,022 genome sequences [60]. In Khailany, et al. study, this variant assumed as the second prevalent mutation after the 8762 C > T [86]. This mutation was founded on 5th Jan 2010 for the first time co-observed with 8782 C > T, and the variants with two these mutations called as S type (S clade) in comparison to the native L clade of the genome [87], leads to improve the replication fidelity in this RNA virus. However, the frequency of two of these mutations has gradually declined over time [87].

The transition mutation of 8782 C > T, in the NSP4 gene, on the other hand, co-appeared with 28144T > C, was assumed as the earlier mutation with decreasing frequency over time [87]. Newer mutations such as 14408 C > T, especially in Europe, lead to a missense mutation in the RdRp gene as P323L, was occurred with a higher frequency for 6319 samples [60]. This mutation is the characteristic of G clade and resulted in higher transmissibility of the virus [87].

The variants containing 23403A > G mutation with an amino acid substitution from aspartic acid to glycine in the S protein, are the other most prevalent variants especially in European samples that created a more transmissible form of SARS-CoV-2 [11]. This mutation is one of the

four prevalent mutations co-occurred in the G clade similarly to the 3037 C > T as a synonymous mutation showed the highest prevalence as above 6300 from over 10,000 samples and was very routine in the Russian samples too [60,88].

Guanidine substitution with Thymidine or Cytidine in the 11083th nucleotide of NSP6 gene was not related to the other frequent mutations and especially related to the European samples [10]. However, in the Koyama study, it was shown that this missense mutation occurred with frequency as about 1100 samples [60].

The other more common point mutation was 26144G > T, in ORF3a resulted in G251V in the proteome level and was more prevalent in European surveyed samples [87]. This type of mutation was assumed as the third prevalent variant after the 8782 C > T and 28144T > C, as a single mutation according to the Yin et al. study [80].

Finally, in one study, a single point mutation has been shown as 21707T > C, corresponds to H49Y amino acid substitution in the S protein, and occurred with a frequency of 0.4% from a total of 4533 surveyed sequences, until 6 May 2020. The first date for creating this variant was estimated as 12 Jan 2020 and emerged again in all the sequences obtained from Mexico. Since then, this mutation has appeared as a singleton in various virus variants worldwide [13]. However, the final effect of this variant has not yet been fully studied [89].

6. Geographic distribution of SARS-CoV-2 mutations

Given that the SARS-CoV-2 originated from China, it can be said that the studied samples from Asia show the least difference compared to the reference genome. However, over time, the average mutation counts per sample increased around the world. This situation can be a sign that the mutant strains are persistent and further spread [87]. From the total 48,635 sequences, in the study of Mercatelli et al., 48,379 samples possessed at least one mutation [82]. On the other hand, it was found that there is no significant difference between continents in the average rate of mutations, but there is a significant difference in the average number of mutations per sample between countries. Actually, among the top 40 countries with the most number of full-length viral genomes, India, Congo, Bangladesh, and Kazakhstan indicated mutations per sample equal to 8.40, 8.30, 9.83, and 9.47, respectively that were the highest values of observed mutations per sample, in comparison to the world average as 7.23 [82]. On the contrary, the number of mutation per sample for sequences originated from European countries such as Germany, Italy, and Greece was calculated as about 5.97, in average, and this value for Japan was reported as 4.55, which was significantly lower than the world's average [82]. These statics for the US, in another study, was about 6.42 with the worldwide average reported as 5.654 [87].

However, in an earlier study, performed in the time interval of Dec

2019-Mar 2020; among a total of 1932 SARS-CoV-2 studied strains, the average mutation counts per sample in American and European populations were much higher than that in Asia populations. Spain, Belgium, and Finland in Europe showed the highest average mutations per sample while Singapore, Japan, and China were on the opposite side [87]. These different results are probably due to the inclusion of the genome information related to China with the lower mutation count to the Asian samples. Actually, because most of the recorded genome sequences in the early months of the outbreak were related to China, this nation showed a lower mutation per sample. Nonetheless, a small number of sequences carrying more than fifty mutations are associated with China, shifting this average rate [90].

As a conclusion of this section, it should be said that according to the investigated samples until Apr 9, 2020, it was shown that there was a strong linear correlation between the number of average mutations per sample and the case fatality rate. Another reported result is the fact that the mutations in ORF1ab, with a significant number of mutations, may contribute the most to the case fatality rate [87].

For comparing the frequency of various clades between different countries, it may be said that in the study of Koyama, it was reported that G clade, was most prevalent in China, while the results of Mercatelli et al. study showed that South America, Europe, Africa, and Oceania had the most frequency of the D614G clade. One reason for this difference is the fact that Mercatelli's study covered the genome sequences until Jun 2020, while the Koyama study is older and the collected samples were from Dec 2019–1 st of May 2020 which most of the samples originated from China, and virus distribution was at this time [90].

The study of Mercatelli showed that the G and GR clades are prevalently present in Europe, while the clade S (L384S mutant) have been mostly seen in the Americas. The L clade, as the reference (basal clade), is mostly represented in Asia samples (predominantly related to China genome sequence). This clade is composed of about 7% of the sequenced genomes until the last day of Mar 2020, and as expected, the G is the most common clade around the world, especially, in North America and Europe. More precisely, North America, especially the US, was reported with the most positive cases of D614G and Q57H, as GH subclade. In South America and Europe, the most frequent clade was GR (containing D614G and RG203KR mutations). For Oceania, on the other hand, there is a balance between all introduced clades. For Africa and Asia, G genomes were with more frequency than other clades. In Europe, Denmark and France, for example, showed the higher presence of GH while the United Kingdom and Portugal show higher numbers of GR. Finally, Russia and Brazil belong to clade GR. Overall, the incidence of G clade is increasing over time, worldwide [90].

G251V, a point mutation as 26144G > T, is a clade first created in Feb 2020. This variant frequently appeared in samples from the United Kingdom, Iceland, the United States, and Australia [60].

As another geographical analysis study in the case of SARS-CoV-2, we can refer to the study of Taboada, et al., surveyed the samples separated from Mexican patients with the Covid-19 positive test. The results of this study showed that all Mexican sequences have 241C > T and 23403A > G, related to G clade [91].

In the Middle East countries, however, there are a few investigations about SARS-CoV-2 mutations. The results of a study that investigated only eleven full-length genome sequences from Lebanon until Mar 2020, showed that the most prevalent variants were 23403A > G (D614G), clade G, and the second common mutant was 14408C > T, in ORF1ab, resulted in a missense mutation as P4715L, one of the three mutations mostly co-occurred with 23403A > G harboring genome. The pattern of variations in this study was equal to the European countries, at the same time which is rational according to their short distances [92].

Another study, for some countries especially from Asia, with data analysis of 2200 full-length genome sequences until Jun 2020. The results of this study established that 241 C > T, 3037 C > T related to the G clade, and 14408 C > T, co-occurred with the mutations in G clade, and also 11083G > T, most prevalent in Europe samples, was the most

prevalent mutations. The other report is this fact that the Iranian samples showed more diversity and interestingly did not reveal 14408C > T mutation, common in most samples. Furthermore, samples from Iran, appeared lower average point mutation frequency than other countries, related to earlier virus epidemics in comparison to the other investigated countries [93]. For other nations, similar variant mapping was showed. However, in the case of Qatar, frame-shift mutations occurred with a higher frequency [93].

The mutations including 3037 C > T, 14408 C > T, 23403A > G, and 241 C > T were more common in Europe and hardly observed in other countries. The American countries, on the other hands, was showed that contain mutations 8782 C > T and 28144T > C. Additionally, the Americas also carried mutations 17747C > T, 17858A > G, and 18060C > T, which all of them showed a frequency more than 0.1% according to the Koyama's study [60,68].

Due to the importance of spike protein in pathogenicity and immunogenicity, we retrieved all Spike sequences submitted to NCBI (until 10th January 2021) for each continent and after deletion of the repeated sequences and multiple alignments, Shannon's entropy plot for each continent was attained by using of BioEdit software (Fig. 2).

7. Conclusion and future perspectives

Natural selection is a major factor in determining the fate of new mutations. The mutations that give a competitive advantage increase. These usually involve the virus escaping from the host's immunity, virus replication, and transmission. Also, those mutations that reduce viral fitness generally tend to eliminate. On the other hand, mutations can occur by chance. Natural selection and chance lead to the evolution of the virus into the hosts.

Whether a mutation changes immune, evasion infectivity, or pathogenicity or some combination of these is yet to be implicit. In some cases, the mutation may transmit with other mutations simultaneously, for example, the mutations that affect the RNA-dependent RNA polymerase, with implications on replication efficiency, proofreading, and the emergence of resistant phenotypes [1]. Finding the answer to these questions depends on worldwide studies to provide comprehensive data on infection and mortality rates as well as the continuous sampling of the genotypes of circulating isolates all around the world. So far, we lack these integrated data and not able to correlate molecular findings with clinical and population-level consequences.

Mutations in the SARS-CoV-2 could have significant effects on the severity of the Covid-19 disease, the contagion, and the virus's stability. It is also responsible for re-infection, vaccination strategy failure, and convalescence plasma and monoclonal antibody (mAb) therapy due to immune-escapes from neutralizing antibody and T cells. It also reduces the quality of diagnostic kits and increases the false negative. Another point in this regard is the emergence of drug-resistant strains.

Some of the important mutations in spike glycoprotein that overshadow the success of vaccines, plasma therapy, and diagnostic kits include L18F, A222V, D614G and Q780E variants in conformational epitopes [97], RBD variants of F338L, S373P and R408T [98,99], S477N and E484A/K mutations [100], two variant amino acids G446V and F456L in a linear epitope [101], and also resistance substitutions at R346, N440, K444, G446, N450, L452, S477, T478, P479, E484, F486, F490, Q493 and P499 [100–103].

By considering the mutation profiles in different parts of the world, and the previous debates about the choice of the vaccine strain, using a single strain for the vaccine can immunize people be to some extent against all strains of SARS-CoV-2. However, in order to achieve maximum vaccine coverage and protection in the human, and also high immune response and neutralizing antibodies (to prevent immune-escape mutants), a vaccine may be prepared in bivalent, trivalent or tetravalent forms. It means it contains a combination of two or three different serotypes. At least one of the serotypes must have a D614G mutation and associated mutations that alter the interaction of

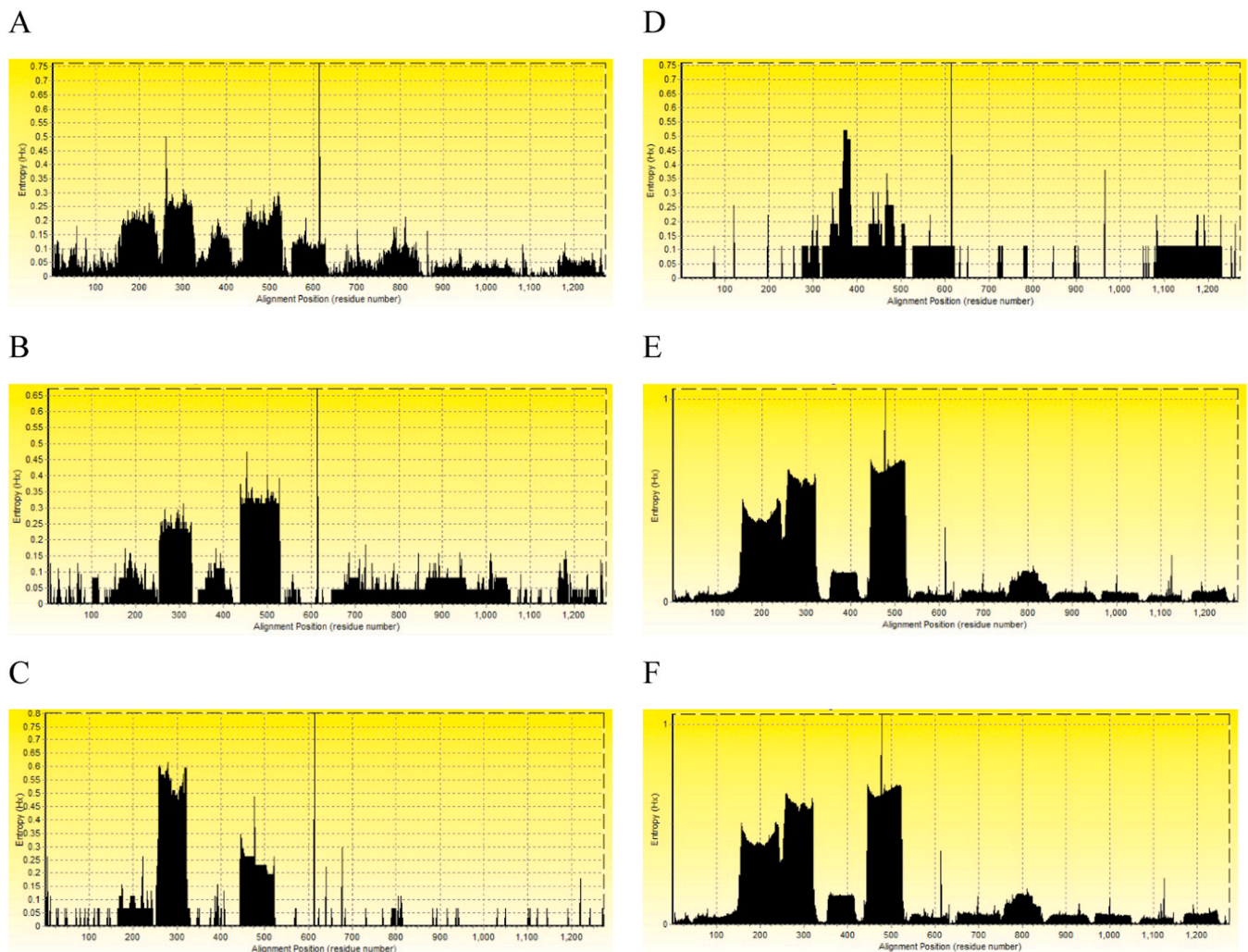


Fig. 2. Entropy plots of all submitted Spike protein sequences. The X-axis shows the positions of amino acids and the Y-axis shows entropy scores for individual positions in Spike protein. (A) Asia, (B) Europe, (C) Africa, (D) South America, (E) North America, and (F) Oceania. Entropy (Hx) values varied from 0 (high conservation) to 1.0 (hypervariable). The amino acids with entropy values less than 0.1 were considered as conserved.

neutralizing antibodies. These mutations change immunogenicity and the structure and sequence of immunodominant linear and conformational epitopes and escape capability from humoral immune pressure, otherwise retain the strong binding affinity toward ACE2 [104,105].

These points are critically crucial for recombinant vaccines such as; mRNAs, DNA, subunits, and viral vector-based vaccines that mostly use only full length S protein or truncated S1 subunit in engineering vaccine constructs. A shift from monovalent to bivalent or trivalent was previously used in the Coronaviridae family members for decades and shows near 100% protection [106,107]. Also, in the killed (inactivated) vaccine, which mainly stimulates the humoral immune system and antibody production, it is necessary to pay attention to these points.

Moreover, due to the high frequency of 14408 (P323L) mutation in RdRp, which increases the possibility of mutations and changes in the entire virus genome (especially in S protein), it is worthwhile to have at least one strain with this mutation in live attenuated and killed vaccine mixture. This strategy enriches the pool and stores the antigen of the vaccine seeds during passages before formulation.

Studies have also shown that in recombinant coronaviruses vaccines, multivalent vaccines for example those containing proteins S, N, and M are more effective than monovalent vaccines that have only S protein [108–110]. In this regard, immunoinformatics and antibody-antigens simulations with wide abilities and areas could accelerate selecting vaccinal strains to incorporate with laboratory tests [106,111].

Furthermore, due to mutations in the cleavage region between S1 and S2, it is better to consider this region in designing recombinant S1 based vaccines that use only S1 length in their construct.

Besides considering the immunologic effects of mutations, the selection of vaccine strains could also assist according to GISAID nomenclature system (<https://www.gisaid.org/>). In the GISAID database, sequenced SARS-CoV-2 genomes were clustered into one of 6 main clades includes G, GH, GR, L, S, and V.

Similarly, most of the points mentioned above are useful for plasma-therapy and therapeutic mAbs. It is due to mutations in the virus at the sites that cause the neutralizing antibodies to fail to inhibit the virus. It decreases the affinity of antibodies to spike glycoproteins [100]. So, it could also recommend that the recovered plasma from individuals must be pooled and then used to reduce the risk of plasma therapy failure (not person to person). Also, in preparing the horse serum for sero-therapy, at least three strains or more should be used for immunization to achieve more strong neutralizing serum, which improves the rate of success of serum therapy.

Using different drugs will cause drug resistance due to mutations. For instance, mutations F480L, V557L, and D484Y in RdRp protein may lead to resistance to Remdesivir [112,113]. Therefore, in the treatment, it may need to use two to three different drugs (combination therapy) against different protein targets to prevent this phenomenon [114,115].

Conflict of interest statement

Authors declare no conflict of interests and/or commercial products or companies.

Acknowledgements

The project was supported by Isfahan University of Medical Sciences [grant number: 198284]. The authors would like to thank infectious Diseases and Tropical Medicine Research Center for kind supports.

References

- [1] M. Pachetti, B. Marini, F. Benedetti, F. Giudici, E. Mauro, P. Storici, C. Masciovecchio, S. Angeletti, M. Ciccozzi, R.C. Gallo, D. Zella, R. Ippodrino, Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant, *J. Transl. Med.* 18 (2020) 1–9.
- [2] R. Sanjuán, P. Domingo-Calap, Mechanisms of viral mutation, *Cell. Mol. Life Sci.* 73 (23) (2016) 4433–4448.
- [3] O.G. Pybus, A.J. Tatem, P. Lemey, Virus evolution and transmission in an ever more connected world, *Proc. R. Soc. B Biol. Sci.* 282 (2015), 20142878.
- [4] M.E. Carter-Timofte, S.E. Jørgensen, M.R. Freytag, M.M. Thomsen, N.-S. B. Andersen, A. Al-Mousawi, et al., Deciphering the role of host genetics in susceptibility to severe COVID-19, *Front. Immunol.* (2020) 11.
- [5] E. Domingo, J. Holland, RNA virus mutations and fitness for survival, *Annu. Rev. Microbiol.* 51 (1) (1997) 151–178.
- [6] E. Domingo, Viruses at the edge of adaptation, *Virology* 270 (2) (2000) 251–253.
- [7] L.A. Holland, E.A. Kaelin, R. Maqsood, B. Estifanos, L.I. Wu, A. Varsani, R. U. Halden, B.G. Hogue, M. Scotch, E.S. Lim, An 81-nucleotide deletion in SARS-CoV-2 ORF7a identified from sentinel surveillance in Arizona (January to March 2020), *J. Virol.* 94 (2020).
- [8] A. Bal, G. Destras, A. Gaymard, M. Bouscambert-Duchamp, M. Valette, Molecular characterization of SARS-CoV-2 in the first COVID-19 cluster in France reveals an amino acid deletion in nsp2 (Asp268del), *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.* 26 (2020) 960–962.
- [9] T. Phan, Genetic diversity and evolution of SARS-CoV-2, *Infect. Genet. Evol. J. Mol. Epidemiol. Evolut. Genet. Infect. Dis.* 81 (2020), 104260.
- [10] S. Liu, J. Shen, S. Fang, K. Li, J. Liu, L. Yang, C.D. Hu, J. Wan, Genetic spectrum and distinct evolution patterns of SARS-CoV-2, *Front. Microbiol.* 11 (2020), 593548.
- [11] M. Islam, M. Hoque, M. Rahman, A.M. Alam ASMRU, J. Puspo, Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity, *Sci. Rep.* 10 (2020) 14004.
- [12] S. Duffy, Why are RNA virus mutation rates so damn high? *PLoS Biol.* 16 (8) (2018), 3000003.
- [13] K.M. Peck, A.S. Lauring, Complexities of viral mutation rates, *J. Virol.* 92 (2018) 14.
- [14] W. Wang, W.-M. Lee, A.G. Mosser, R.R. Rueckert, WIN 52035-dependent human rhinovirus 16: assembly deficiency caused by mutations near the canyon surface, *J. Virol.* 72 (2) (1998) 1210–1218.
- [15] B.A. Heinz, R.R. Rueckert, D.A. Shepard, F.J. Dutko, M.A. McKinlay, M. Fancher, M.G. Rossmann, J. Badger, T.J. Smith, Genetic and molecular analyses of spontaneous mutants of human rhinovirus 14 that are resistant to an antiviral compound, *J. Virol.* 63 (6) (1989) 2476–2485.
- [16] J.C. De La Torre, C. Giachetti, B.L. Semler, J.J. Holland, High frequency of single-base transitions and extreme frequency of precise multiple-base reversion mutations in poliovirus, *Proc. Natl. Acad. Sci.* 89 (7) (1992) 2531–2535.
- [17] J.W. Drake, J.J. Holland, Mutation rates among RNA viruses, *Proc. Natl. Acad. Sci.* 96 (24) (1999) 13910–13913.
- [18] J. De la Torre, E. Wimmer, J. Holland, Very high frequency of reversion to guanidine resistance in clonal pools of guanidine-dependent type 1 poliovirus, *J. Virol.* 64 (2) (1990) 664–671.
- [19] H.-W. Huang, P.-H. Chu, C.-H. Pan, C.-F. Wang, C.-C. Lin, P.-L. Lu, Y.S. Chen, Y. Shi, H.J. Su, L.C. Chou, Y.Y. Lin, H.F. Lee, B.C. Chen, T.S. Huang, Y.C. Tyan, C. H. Chuang, Y.C. Yen, P.Y. Chu, Evolutionary histories of coxsackievirus B5 and swine vesicular disease virus reconstructed by phylodynamic and sequence variation analyses, *Sci. Rep.* 8 (1) (2018) 1–12.
- [20] J.M. Cuevas, M. Combe, M. Torres-Puente, R. Garijo, S. Guix, J. Buesa, J. Rodríguez-Díaz, R. Sanjuán, Human norovirus hyper-mutation revealed by ultra-deep sequencing, *Infect. Genet. Evol.* 41 (2016) 233–239.
- [21] J.M. Cuevas, F. González-Candelas, A. Moya, R. Sanjuán, Effect of ribavirin on the mutation rate and spectrum of hepatitis C virus in vivo, *J. Virol.* 83 (11) (2009) 5760–5764.
- [22] R.M. Ribeiro, H. Li, S. Wang, M.B. Stoddard, G.H. Learn, B.T. Korber, T. Bhattacharya, J. Guedj, E.H. Parrish, B.H. Hahn, G.M. Shaw, A.S. Perelson, Quantifying the diversification of hepatitis C virus (HCV) during primary infection: estimates of the in vivo mutation rate, *PLoS Pathog.* 8 (8) (2012), 1002881.
- [23] J.H.-O. Pettersson, V. Eldholm, S.J. Seligman, Å. Lundkvist, A.K. Falconar, M. W. Gaunt, D. Musso, A. Nougairède, R. Charrel, E.A. Gould, X. de Lamballerie, How did Zika virus emerge in the Pacific Islands and Latin America? *MBio* 7 (2016) 5.
- [24] C. Koo, W.P. Tien, H. Xu, J. Ong, J. Rajarethinam, Y.L. Lai, L.C. Ng, H. C. Hapuarachchi, Highly selective transmission success of dengue virus type 1 lineages in a dynamic virus population: an evolutionary and fitness perspective, *iScience* 6 (2018) 38–51.
- [25] C.-C. Hon, T.-Y. Lam, Z.-L. Shi, A.J. Drummond, C.-W. Yip, F. Zeng, P.Y. Lam, F. C. Leung, Evidence of the recombinant origin of a bat severe acute respiratory syndrome (SARS)-like coronavirus and its implications on the direct ancestor of SARS coronavirus, *J. Virol.* 82 (4) (2008) 1819–1826.
- [26] K. Pyrc, R. Dijkman, L. Deng, M.F. Jebbink, H.A. Ross, B. Berkhout, L. van der Hoek, Mosaic structure of human coronavirus NL63, one thousand years of evolution, *J. Mol. Biol.* 364 (5) (2006) 964–973.
- [27] E.H. Lau, C.A. Hsiung, B.J. Cowling, C.-H. Chen, L.-M. Ho, T. Tsang, C.W. Chang, C.A. Donnelly, G.M. Leung, A comparative epidemiologic analysis of SARS in Hong Kong, Beijing and Taiwan, *BMC Infect. Dis.* 10 (1) (2010) 1–9.
- [28] M. Combe, R. Sanjuán, Variation in RNA virus mutation rates across host cells, *PLoS Pathog.* 10 (1) (2014), 1003855.
- [29] P. Davis, A. Rambaut, H. Bourhy, E. Holmes, The evolutionary dynamics of canid and mongoose rabies virus in Southern Africa, *Arch. Virol.* 152 (7) (2007) 1251–1258.
- [30] J.D. Parvin, A. Moscona, W. Pan, J. Leider, P. Palese, Measurement of the mutation rates of animal viruses: influenza A virus and poliovirus type 1, *J. Virol.* 59 (2) (1986) 377–383.
- [31] E. Nobusawa, K. Sato, Comparison of the mutation rates of human influenza A and B viruses, *J. Virol.* 80 (7) (2006) 3675–3678.
- [32] C.-F. Yang, C.K. Wang, S.J. Tollefson, R. Piyaaratna, L.D. Lintao, M. Chu, A. Liem, M. Mark, R.R. Spaete, Jr Crowe JE, J.V. Williams, Genetic diversity and evolution of human metapneumovirus fusion protein over twenty years, *Virol. J.* 6 (1) (2009) 138.
- [33] L.W. Pomeroy, O.N. Bjørnstad, E.C. Holmes, The evolutionary and epidemiological dynamics of the paramyxoviridae, *J. Mol. Evol.* 66 (2) (2008) 98–106.
- [34] X. Zhang, L.J. Rennick, W.P. Duprex, B.K. Rima, Determination of spontaneous mutation frequencies in measles virus under nonselective conditions, *J. Virol.* 87 (5) (2013) 2686–2692.
- [35] S. Van Niekerk, M. Venter, Replacement of previously circulating respiratory syncytial virus subtype B strains with the BA genotype in South Africa, *J. Virol.* 85 (17) (2011) 8789–8797.
- [36] S.A. Carroll, B.H. Bird, P.E. Rollin, S.T. Nichol, Ancient common ancestry of Crimean-Congo hemorrhagic fever virus, *Mol. Phylogenet. Evol.* 55 (3) (2010) 1103–1110.
- [37] J. Matthijnsens, M. Rahman, M. Ciarlet, M. Zeller, E. Heylen, T. Nakagomi, R. Uchida, Z. Hassan, T. Azim, O. Nakagomi, M. Van Ranst, Reassortment of human rotavirus gene segments into G11 rotavirus strains, *Emerg. Infect. Dis.* 16 (4) (2010) 625–630.
- [38] J.M. Leider, P. Palese, F.I. Smith, Determination of the mutation rate of a retrovirus, *J. Virol.* 62 (9) (1988) 3084–3091.
- [39] L.M. Mansky, In vivo analysis of human T-cell leukemia virus type 1 reverse transcription accuracy, *J. Virol.* 74 (20) (2000) 9525–9531.
- [40] L.M. Mansky, S. Preveral, L. Selig, R. Benarous, S. Benichou, The interaction of vpr with uracil DNA glycosylase modulates the human immunodeficiency virus type 1 In vivo mutation rate, *J. Virol.* 74 (15) (2000) 7039–7047.
- [41] F. Gao, Y. Chen, D.N. Levy, J.A. Conway, T.B. Kepler, H. Hui, Unselected mutations in the human immunodeficiency virus type 1 genome are mostly nonsynonymous and often deleterious, *J. Virol.* 78 (5) (2004) 2426–2433.
- [42] A. Parsyan, C. Szmaragd, J.-P. Allain, D. Candotti, Identification and genetic diversity of two human parvovirus B19 genotype 3 subtypes, *J. Gen. Virol.* 88 (2) (2007) 428–431.
- [43] L.A. Shackleton, A. Rambaut, O.G. Pybus, E.C. Holmes, JC virus evolution and its association with human populations, *J. Virol.* 80 (20) (2006) 9928–9933.
- [44] J.W. Drake, C.B. Hwang, On the mutation rate of herpes simplex virus type 1, *Genetics* 170 (2) (2005) 969–970.
- [45] J. Jiricny, Postreplicative mismatch repair, *Cold Spring Harb. Perspect. Biol.* 5 (4) (2013), 012633.
- [46] F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.W. Tao, J. H. Tian, Y.Y. Pei, M.L. Yuan, Y.L. Zhang, F.H. Dai, Y. Liu, Q.M. Wang, J.J. Zheng, L. Xu, E.C. Holmes, Y.Z. Zhang, A new coronavirus associated with human respiratory disease in China, *Nature* 579 (7798) (2020) 265–269.
- [47] D.A. Benson, I. Karsch-Mizrachi, D.J. Lipman, J. Ostell, B.A. Rapp, D.L. Wheeler, GenBank, *Nucleic Acids Res.* 28 (1) (2000) 15–18.
- [48] L. Mousavizadeh, S. Ghasemi, Genotype and phenotype of COVID-19: their roles in pathogenesis, *J. Microbiol. Immunol. Infect.* (2020).
- [49] K. Schubert, E.D. Karousis, A. Jomaa, A. Scialoja, B. Echeverria, L.-A. Gurzeler, M. Leibundgut, V. Thiel, O. Mühlemann, N. Ban, SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation, *Nat. Struct. Mol. Biol.* 27 (10) (2020) 959–966.
- [50] X. Lei, X. Dong, R. Ma, W. Wang, X. Xiao, Z. Tian, C. Wang, Y. Wang, L. Li, L. Ren, F. Guo, Z. Zhao, Z. Zhou, Z. Xiang, J. Wang, Activation and evasion of type I interferon responses by SARS-CoV-2, *Nat. Commun.* 11 (1) (2020) 1–12.
- [51] T. Klemm, G. Ebert, D.J. Calleja, C.C. Allison, L.W. Richardson, J.P. Bernardini, B. G. Lu, N.W. Kuchel, C. Grohmann, Y. Shibata, Z.Y. Gan, J.P. Cooney, M. Doerflinger, A.E. Au, T.R. Blackmore, G.J. van der Heden van Noort, P. P. Geurink, H. Ovaa, J. Newman, A. Riboldi-Tunnicliffe, P.E. Czabotar, J. P. Mitchell, R. Feltham, B.C. Lechtenberg, K.N. Lowes, G. Dewson, M. Pellegrini, G. Lessene, D. Komander, Mechanism and inhibition of the papain-like protease, PLpro, of SARS-CoV-2, *EMBO J.* 39 (18) (2020), 106275.

- [52] N. Le Bert, A.T. Tan, K. Kunasegaran, C. Tham, M. Hafezi, A. Chia, M. Chng, M. Lin, N. Tan, M. Linster, W.N. Chia, M.I. Chen, L.F. Wang, E.E. Ooi, S. Kalimuddin, P.A. Tambyah, J.G. Low, Y.J. Tan, A. Bertoletti, SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls, *Nature* 584 (7821) (2020) 457–462.
- [53] Y. Huang, C. Yang, X.-f Xu, W. Xu, S.-w Liu, Structural and functional properties of SARS-CoV-2 spike protein: potential antiviral drug development for COVID-19, *Acta Pharmacol. Sin.* 41 (9) (2020) 1141–1149.
- [54] Y. Ren, T. Shu, D. Wu, J. Mu, C. Wang, M. Huang, Y. Han, X.Y. Zhang, W. Zhou, Y. Qiu, J. Zhou, The ORF3a protein of SARS-CoV-2 induces apoptosis in cells, *Cell. Mol. Immunol.* 17 (8) (2020) 881–883.
- [55] F. De Maio, E.L. Cascio, G. Babini, M. Sali, S. Della Longa, B. Tilocca, et al. Enhanced binding of SARS-CoV-2 Envelope protein to tight junction-associated PALS1 could play a key role in COVID-19 pathogenesis. 2020.
- [56] S. Srinivasan, H. Cui, Z. Gao, M. Liu, S. Lu, W. Mkandawire, O. Narykov, M. Sun, D. Korkin, Structural genomics of SARS-CoV-2 indicates evolutionary conserved functional regions of viral proteins, *Viruses* 12 (4) (2020) 360.
- [57] H.-w Jiang, H.-n Zhang, Q.-f Meng, J. Xie, Y. Li, H. Chen, Y.X. Zheng, X.N. Wang, H. Qi, J. Zhang, P.H. Wang, Z.G. Han, S.C. Tao, SARS-CoV-2 Orf9b suppresses type I interferon responses by targeting TOM70, *Cell. Mol. Immunol.* 17 (9) (2020) 998–1000.
- [58] A. Gómez-Carballa, X. Bello, J. Pardo-Seco, F. Martínón-Torres, A. Salas, Mapping genome variation of SARS-CoV-2 worldwide highlights the impact of COVID-19 super-spreaders, *Genome Res.* 30 (10) (2020) 1434–1448.
- [59] Y. Toyoshima, K. Nemoto, S. Matsumoto, Y. Nakamura, K. Kiyotani, SARS-CoV-2 genomic variations associated with mortality rate of COVID-19, *J. Hum. Genet.* 65 (2020) 1075–1082.
- [60] T. Koyama, D. Platt, L. Parida, Variant analysis of SARS-CoV-2 genomes, *Bull. World Health Organ.* 98 (7) (2020) 495–504.
- [61] P. Forster, L. Forster, C. Renfrew, M. Forster, Phylogenetic network analysis of SARS-CoV-2 genomes, *Proc. Natl. Acad. Sci.* 117 (17) (2020) 9241–9243.
- [62] Y. Toyoshima, K. Nemoto, S. Matsumoto, Y. Nakamura, K. Kiyotani, SARS-CoV-2 genomic variations associated with mortality rate of COVID-19, *J. Hum. Genet.* 65 (12) (2020) 1075–1082.
- [63] N.K. Yellapu, S. Patel, B. Zhang, R. Meier, L. Neums, D. Pei, Q. Xia, D. Rotich, R. C. Zimmermann, E. Nissen, S. Bell-Glenn, W. Shae, J. Hu, P. Chalise, L. Chollet-Hinton, D.C. Koestler, J.A. Thompson, Evolutionary analysis of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) reveals genomic divergence with implications for universal vaccine efficacy, *Vaccines* 8 (4) (2020) 591.
- [64] Z. Zhao, B.A. Sokhansanj, C. Malhotra, K. Zheng, G.L. Rosen, Genetic grouping of SARS-CoV-2 coronavirus sequences using informative subtype markers for pandemic spread visualization, *PLoS Comput. Biol.* 16 (9) (2020), 1008269.
- [65] M. Worobey, J. Pekar, B.B. Larsen, M.I. Nelson, V. Hill, J.B. Joy, A. Rambaut, M. A. Suchard, J.O. Wertheim, P. Lemey, The emergence of SARS-CoV-2 in Europe and North America, *Science* 370 (6516) (2020) 564–570.
- [66] I. Karacan, T.K. Akgun, N.B. Agaoglu, A. Irvem, G. Alkurt, J. Yildiz, B. Kose, A. S. Ozel, L.N. Altunal, N.D. Can, Y.K. Demirkol, M. Aydin, O.A. Dogan, L. Doganay, G.D. Doganay, The origin of SARS-CoV-2 in Istanbul: Sequencing findings from the epicenter of the pandemic in Turkey, *North. Clin. Istanbul* 7 (3) (2020) 203–209.
- [67] Z. Fattahi, M. Mohseni, K. Jalalvand, F.A. Moghadam, A. Ghaziasadi, F. Keshavarzi, et al. Two independent introductions of SARS-CoV-2 into the Iranian outbreak. *medRxiv*. 2020:2020.11.16.20229047.
- [68] B. Korbner, W.M. Fischer, S. Gnanakaran, H. Yoon, J. Theiler, W. Abfalterer, N. Hengartner, E.E. Giorgi, T. Bhattacharya, B. Foley, K.M. Hastie, M.D. Parker, D.G. Partridge, C.M. Evans, T.M. Freeman, T.I. de Silva, G. Sheffield COVID-Genomics, C. McDaniel, L.G. Perez, H. Tang, A. Moon-Walker, S.P. Whelan, C. C. LaBranche, E.O. Saphire, D.C. Montefiori, Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus, *Cell* 182 (4) (2020) 812–827. e19.
- [69] E. Volz, V. Hill, J.T. McCrone, A. Price, D. Jorgensen, Á. O'toole, J. Southgate, R. Johnson, B. Jackson, F.F. Nascimento, S.M. Rey, S.M. Nicholls, R. M. Colquhoun, A. da Silva Filipe, J. Shepherd, D.J. Pascall, R. Shah, N. Jesudason, K. Li, R. Jarrett, N. Pacchiarini, M. Bull, L. Geidelberg, I. Siveroni, C. COG-UK, I. Goodfellow, N.J. Loman, O.G. Pybus, D.L. Robertson, E.C. Thomson, A. Rambaut, T.R. Connor, Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity, *Cell* 184 (1) (2021) 64–75.e11.
- [70] S.-W. Huang, S.O. Miller, C.-H. Yen, S.-F. Wang, Impact of genetic variability in ACE2 expression on the evolutionary dynamics of SARS-CoV-2 spike D614G mutation, *Genes* 12 (2021) 1.
- [71] C. Wu, Y. Liu, Y. Yang, P. Zhang, W. Zhong, Y. Wang, Q. Wang, Y. Xu, M. Li, X. Li, M. Zheng, L. Chen, H. Li, Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods, *Acta Pharm. Sin. B* 10 (2020) 766–788.
- [72] Organization PAHOWH. Occurrence of variants of SARS-CoV-2 in the Americas Washington, D.C.: PAHO/WHO; 202126 January 2021.
- [73] S. Fiorentini, S. Messali, A. Zani, F. Caccuri, M. Giovanetti, M. Ciccozzi, A. Caruso, First detection of SARS-CoV-2 spike protein N501 mutation in Italy in August, 2020, *Lancet Infect. Dis.* (2021).
- [74] C. Yi, X. Sun, J. Ye, L. Ding, M. Liu, Z. Yang, X. Lu, Y. Zhang, L. Ma, W. Gu, A. Qu, J. Xu, Z. Shi, Z. Ling, B. Sun, Key residues of the receptor binding motif in the spike protein of SARS-CoV-2 that interact with ACE2 and neutralizing antibodies, *Cell. Mol. Immunol.* 17 (6) (2020) 621–630.
- [75] E. Mahase, Covid-19: What New Variants are Emerging And How are They Being Investigated? British Medical Journal Publishing Group, 2021.
- [76] C.M. Voloch, R. da Silva Francisco, L.G.P. de Almeida, C.C. Cardoso, O. J. Brustolini, A.L. Gerber, A.P.C. Guimarães, D. Mariani, R.M. da Costa, O. C. Ferreira, T.S. Frauches, C.M.B. de Mello, I.C. Leitão, R.M. Galliez, D.S. Paffe, T. M.P.P. Castineiras, A. Tanuri, A.T.R. de Vasconcelos, Genomic characterization of a novel SARS-CoV-2 lineage from Rio de Janeiro, Brazil, *J. Virol.* (2021).
- [77] N.R. Faria, I.M. Claro, D. Candido, L. Moyses Franco, P.S. Andrade, T.M. Coletti, et al., Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings, *Virological* (2021).
- [78] H. Tegally, E. Wilkinson, M. Giovanetti, A. Iranzadeh, V. Fonseca, J. Giandhari, et al. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv*. 2020.
- [79] J. Zimmer, CaC. Coronavirus variants and mutations The New York Times 2021.
- [80] C. Yin, Genotyping coronavirus SARS-CoV-2: methods and implications, *Genomics* (2020).
- [81] C. Wang, Z. Liu, Z. Chen, X. Huang, M. Xu, T. He, Z. Zhang, The establishment of reference sequence for SARS-CoV-2 and variation analysis, *J. Med. Virol.* 92 (6) (2020) 667–674.
- [82] D. Mercatelli, F.M. Giorgi, Geographic and genomic distribution of SARS-CoV-2 mutations, *Front. Microbiol.* (2020).
- [83] E.C. Rouchka, J.H. Chariker, D. Chung, Variant analysis of 1,040 SARS-CoV-2 genomes, *PLoS One* 15 (11) (2020), 0241535.
- [84] M.R. Islam, M.N. Hoque, M.S. Rahman, A.S.M.R.U. Alam, M. Akther, J.A. Puspo, S. Akter, M. Sultana, K.A. Crandall, M.A. Hossain, Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity, *Sci. Rep.* 10 (1) (2020) 1–9.
- [85] M. Wang, M. Li, R. Ren, A. Brave, S. van der Werf, E.-Q. Chen, et al. International expansion of a novel SARS-CoV-2 mutant. *medRxiv*. 2020.
- [86] R.A. Khailany, M. Safdar, M. Ozaslan, Genomic characterization of a novel SARS-CoV-2, *Gene Rep.* 19 (2020), 100682.
- [87] H.-C. Yang, C.-h Chen, J.-H. Wang, H.-C. Liao, C.-T. Yang, C.-W. Chen, et al. Genomic, geographic and temporal distributions of SARS-CoV-2 mutations. *bioRxiv*. 2020.
- [88] L. Kozlovskaya, A. Piniaeva, G. Ignatyev, A. Selivanov, A. Shishova, A. Kovpak, I. Gordeychuk, Y. Ivin, A. Berestovskaya, E. Prokhortchouk, D. Protsenko, M. Rychev, A. Ishmukhametov, Isolation and phylogenetic analysis of SARS-CoV-2 variants collected in Russia during the COVID-19 outbreak, *Int. J. Infect. Dis.* 99 (2020) 40–46.
- [89] M. Gui, W. Song, H. Zhou, J. Xu, S. Chen, Y. Xiang, X. Wang, Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding, *Cell Res.* 27 (1) (2017) 119–129.
- [90] P. Manickavasagam, Spike Protein of SARS-CoV-2: Impact of Single Amino Acid Mutation and Effect of Drug Binding to the Variant-in Silico Analysis. 2020.
- [91] B. Taboada, J.A. Vázquez-Pérez, J.E. Muñoz-Medina, P. Ramos-Cervantes, M. Escalera-Zamudio, C. Boukadida, et al. Genomic analysis of early SARS-CoV-2 strains introduced in Mexico. *bioRxiv*. 2020.
- [92] M. Abou-Hamdan, K. Hamze, A. Abdel Sater, H. Akl, N. El-Zein, I. Dandache, F. Abdel-Sater, Variant analysis of the first Lebanese SARS-CoV-2 isolates, *Genomics* 113 (2021) 892–895.
- [93] K.M. Bindayna, S. Crinion, Variant analysis of SARS-CoV-2 genomes in the Middle East. *bioRxiv*. 2020.
- [94] F. Fratev, The N501Y and K417N mutations in the spike protein of SARS-CoV-2 alter the interactions with both hACE2 and human derived antibody: A Free energy of perturbation study. *bioRxiv*. 2020.
- [95] A. Lopez-Rincon, C. Perez-Romero, A. Tonda, L. Mendoza-Maldonado, E. Claassen, J. Garssen, et al. Design of Specific Primer Sets for the Detection of B. 1.1.7, B. 1.351 and P. 1 SARS-CoV-2 Variants using Deep Learning. *bioRxiv*. 2021.
- [96] P.A.G. Ferrareze, V.B. Franceschi, A. de Menezes Mayer, G.D. Caldana, R.A. Zimerman, C.E. Thompson, E484K as an innovative phylogenetic event for viral evolution: Genomic analysis of the E484K spike mutation in SARS-CoV-2 lineages from Brazil. *BioRxiv*. 2021.
- [97] B.-z Zhang, Y.-f Hu, L.-l Chen, T. Yau, Y.-g Tong, J.-c Hu, J.P. Cai, K.H. Chan, Y. Dou, J. Deng, X.L. Wang, I.F. Hung, K.K. To, K.Y. Yuen, J.D. Huang, Mining of epitopes on spike protein of SARS-CoV-2 from COVID-19 patients, *Cell Res.* 30 (8) (2020) 702–704.
- [98] M. Yuan, N.C. Wu, X. Zhu, C.D. Lee, R. So, H. Lv, C. Mok, I.A. Wilson, A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV, *Science* 368 (6491) (2020) 630–633.
- [99] J. Huo, Y. Zhao, J. Ren, D. Zhou, H. Duyvesteyn, H. Ginn, et al., Neutralization of SARS-CoV-2 by destruction of the prefusion spike, *Cell Host Microbe* (2020).
- [100] Z. Liu, L.A. VanBlargan, P.W. Rothlauf, L.-M. Bloyet, R.E. Chen, S. Stumpf, et al. Landscape analysis of escape variants identifies SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. Available at SSRN 3725763. 2020.
- [101] T. Li, X. Han, Y. Wang, C. Gu, J. Wang, C. Hu, et al. A key linear epitope for a potent neutralizing antibody to SARS-CoV-2 S-RBD. *bioRxiv*. 2020.
- [102] A. Baum, B.O. Fulton, E. Wloga, R. Copin, K.E. Pascal, V. Russo, S. Giordano, K. Lanza, N. Negron, M. Ni, Y. Wei, G.S. Atwal, A.J. Murphy, N. Stahl, G. D. Yancopoulos, C.A. Kyrtatos, Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies, *Science* 369 (6506) (2020) 1014–1018.
- [103] Y. Weisblum, F. Schmidt, F. Zhang, J. DaSilva, D. Poston, J.C. Lorenzi, F. Muecksch, M. Rutkowska, H.H. Hoffmann, E. Michailidis, C. Gaebler, M. Agudelo, A. Cho, Z. Wang, A. Gazumyan, M. Cipolla, L. Luchsinger, C. D. Hillyer, M. Caskey, D.F. Robbani, C.M. Rice, M.C. Nussenzweig,

- T. Hatzioannou, P.D. Bieniasz, Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants, *Elife* 9 (2020), e61312.
- [104] K. Saylor, F. Gillam, T. Lohneis, C. Zhang, Designs of antigen structure and composition for improved protein-based vaccine efficacy, *Front. Immunol.* 11 (2020) 283.
- [105] P.K. Singh, U. Kulsum, S.B. Rufai, S.R. Mudliar, S. Singh, Mutations in SARS-CoV-2 leading to antigenic variations in spike protein: a challenge in vaccine development, *J. Lab. Physicians* 12 (2) (2020) 154–160.
- [106] M. Vasfi Marandi, M. Malekan, M. Ranjbar, N. Dadashpour Davachi, S. Alamian, Sequencing and In Silico Multi-aspect Analysis of S1 Glycoprotein in 793/B Serotype of Infectious Bronchitis Virus Isolated From Iran in 2003 and 2011, *Arch. Razi Inst.* 73 (3) (2018) 183–198.
- [107] P. Box, H. Holmes, P. Finney, R. Froymann, Infectious bronchitis in laying hens: the relationship between haemagglutination inhibition antibody levels and resistance to experimental challenge, *Avian Pathol.* 17 (2) (1988) 349–361.
- [108] L. Tian, H.-n Wang, D. Lu, Y.-f Zhang, T. Wang, R.-m Kang, The immunoreactivity of a chimeric multi-epitope DNA vaccine against IBV in chickens, *Biochem. Biophys. Res. Commun.* 377 (1) (2008) 221–225.
- [109] T. Yang, H.-N. Wang, X. Wang, J.-N. Tang, R. Gao, J. Li, Z.C. Guo, Y.L. Li, Multivalent DNA vaccine enhanced protection efficacy against infectious bronchitis virus in chickens, *J. Vet. Med. Sci.* 71 (12) (2009) 1585–1590.
- [110] H. Jiao, Z. Pan, Y. Yin, S. Geng, L. Sun, X. Jiao, Oral and nasal DNA vaccines delivered by attenuated *Salmonella enterica* serovar Typhimurium induce a protective immune response against infectious bronchitis in chickens, *Clin. Vaccin. Immunol.* 18 (7) (2011) 1041–1045.
- [111] M.M. Ranjbar, M.M. Ebrahimi, S. Shahsavandi, T. Farhadi, A. Mirjalili, M. Tebianian, M.H. Motedayen, Novel applications of immuno-bioinformatics in vaccine and bio-product developments at research institutes, *Arch. Razi Inst.* 74 (3) (2019) 219–233.
- [112] M. Martinot, A. Jary, S. Fafi-Kremer, V. Leducq, H. Delagrevier, M. Garnier, J. Pacanowski, A. Mékinian, F. Pirenne, P. Tiberghien, V. Calvez, C. Humbrecht, A.G. Marcelin, K. Lacombe, Remdesivir failure with SARS-CoV-2 RNA-dependent RNA-polymerase mutation in a B-cell immunodeficient patient with protracted Covid-19, *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* (2020).
- [113] M.L. Agostini, E.L. Andres, A.C. Sims, R.L. Graham, T.P. Sheahan, X. Lu, E. C. Smith, J.B. Case, J.Y. Feng, R. Jordan, A.S. Ray, T. Cihlar, D. Siegel, R. L. Mackman, M.O. Clarke, R.S. Baric, M.R. Denison, Coronavirus susceptibility to the antiviral remdesivir (GS-5734) is mediated by the viral polymerase and the proofreading exoribonuclease, *MBio* 9 (2018) 2.
- [114] L. Menéndez-Arias, D.D. Richman, Editorial overview: antivirals and resistance: advances and challenges ahead, *Curr. Opin. Virol.* 8 (2014) 8.
- [115] S. Mason, J.P. Devincenzo, S. Toovey, J.Z. Wu, R.J. Whitley, Comparison of antiviral resistance across acute and chronic viral infections, *Antivir. Res.* 158 (2018) 103–112.